

## **A METHOD FOR OBTAINING IMPROVED FERTILITY RESTORER LINES FOR TRANSGENIC MALE STERILE CROP PLANTS AND A DNA CONSTRUCT FOR USE IN SAID METHOD**

### **5 FIELD OF INVENTION**

The present invention relates to a method for obtaining improved fertility restorer lines for male sterile crop plants developed using transgenic approaches for hybrid seed production and a DNA construct for use in said method. In particular, the present invention relates to a method for stable, enhanced and extended temporal expression of a transgene and a method for the development of improved fertility restorer lines for hybrid seed production using molecular approaches in crop plants. The invention relates to the simultaneous use of two different gene sequences encoding the same protein product, one being the naturally occurring wild type sequence and the other sequence generated by modification of the wild type sequence for expression in crop plants by using codon degeneracy to avoid homology between the two sequences at the DNA and mRNA levels, each of the said sequences being placed under independent transcriptional control of different overlapping plant tissue-specific regulatory elements in the same DNA construct.

### **BACKGROUND OF INVENTION**

Genetic transformation of crop plants with heterologous genes is widely recognized as an effective strategy for the introduction of desirable traits such as male sterility/restoration for hybrid seed production, insect resistance, nutritional enhancement, resistance to biotic and abiotic stresses, etc. in crop plants. Expression levels of the transgene(s) are critical for achieving the desired phenotype in transgenic plants. Transgene expression levels can be substantially enhanced by modifications at the transcriptional level (by the use of strong promoters/enhancers) or at the translational level (by modulating parameters viz., codon usage, mRNA stability, etc. associated with coding regions of transgenes). However, the existence of various regulatory mechanisms in plants renders the introduced transgene(s) susceptible to down-regulation which, in many instances, also leads to transgene silencing. Homology based gene silencing (HBGS) has been reported to occur extensively in transgenic plants (Meyer and Saedler 1996, Ann. Rev. Plant Physiol. Plant Mol. Biol. 47:23-48; Vaucheret and Fagard 2001, Trends Genet. 17:29-35) and can occur at the transcriptional (transcriptional gene silencing, TGS) as well as at the post transcriptional levels (post transcriptional gene silencing, PTGS). Such

situations can arise when (i) multiple copies of a gene cassette integrate into plants during transformation (van der Krol *et al.* 1990, Plant Cell 2:291-299), (ii) introduced transgenes are transcribed using homologous promoters (Mol *et al.* 1989, Plant Mol. Biol. 13:287-294) or (iii) an introduced gene has homology with an endogenous gene in the coding region (Napoli *et al.* 1990, Plant Cell 2:279-289). Several mechanisms have been suggested to explain the phenomena of HBGS (Matzke and Matzke 1995, Plant Physiol. 107:679-685; Meyer and Saedler 1996, Ann. Rev. Plant Physiol Plant Mol. Biol. 47:23-48; Fire 1999, Trends Genet. 15:358-363; Hamilton and Baulcombe 1999, Science 286:950-952; Steimer *et al.* 2000, Plant Cell 12:1165-1178), the common denominator being that the homology triggers cellular recognition mechanisms that result in silencing of the repeated genes. The present invention relates to a method for enhancing expression levels of transgene(s) while reducing its susceptibility to homology based post transcriptional gene silencing. The strategy of the present invention has been tested by enhancing expression of a restorer gene (*barstar*) to facilitate effective restoration of male fertility in *barnase*-containing male sterile transgenic plants for hybrid seed production in crop plants.

The contributions of hybrids towards enhancing crop productivity are well recognized. Increase in crop productivity is due to the phenomenon of heterosis or hybrid vigor in which F1 hybrid plants generated by crosses between two genetically diverse parents exhibit an improved yield than either of the two parents (Banga 1992, *In* Breeding Oilseed Brassicas Eds. Labana KS, Banga SS and Banga SK, Narosa Publishing House, New Delhi; Pradhan *et al.* 1993, Euphytica 69:219-229). Cross pollination is essential for the production of hybrids. In normally self-pollinating crop plants (for example, *Brassica* sp., rice and wheat), one of the parents needs to be male sterile to facilitate cross pollination. In crop plants wherein seeds are the desired economic products, availability of a suitable restorer system in the male parent is essential in order to achieve seed set in the F1 hybrids.

One of the approaches for generation of male sterile/restorer lines for hybrid seed production is the use of cytoplasmic male sterility (CMS)/restorer systems. CMS is a maternally inherited phenomenon, the genetic determinants of which are located in genomes of the cytoplasmic organelles, the mitochondria. Such plants are severely impaired in their ability to produce functional pollen grains. Restorer genes for CMS systems are dominant nuclear genes that suppress male sterile effects of the cytoplasm (i.e., mitochondria). On being incorporated into the male parent, they can be used as

restorers of male fertility in F1 hybrids. CMS systems have been widely used in the production of hybrids in sorghum, sunflower, pearl millet and sugar beet. However, their use has been limited in corn, wheat and oilseed Brassicas due to linkage of undesirable traits such as increased disease susceptibility, chlorosis, distortion of petals, poor nectary function, etc. with CMS in these systems (McVetty *et al* 1989, Can J. Plant Sci. 69:915-918; Burns *et al* 1991, Can. J. Plant Sci. 71:655-661; Williams 1995, Trends Biotech. 13:344-349).

With the advent of recombinant DNA and plant transformation technologies, genetic engineering of male sterility and its restoration have emerged as tangible options for the development of male sterile and restorer lines for hybrid seed production. Male sterility can be artificially induced in plants by introducing a male sterility gene comprising a DNA sequence coding, for example, a cytotoxic product. The cytotoxic product may be a lethal gene under transcriptional control of a promoter which is predominantly active in selective tissue(s) of male reproductive organs in plants. Tapetum, which forms the innermost layers of the anther wall, is one of the most important tissues associated with pollen development. Disruption of tapetal cells by the expression of toxic proteins would therefore impair pollen development leading to male sterile plants. Male sterility was successfully induced in transgenic tobacco and oilseed rape (*Brassica napus*) by targeted expression of a ribonuclease [*barnase* from *Bacillus amyloliquefaciens* (Hartley 1989, Trends Biochem. Sciences 14:450-454) or *Rnase T1* from *Aspergillus oryzae*] in the tapetal tissues using a tapetum-specific promoter (TA29) from tobacco (Mariani *et al* 1990, Nature 347:737-741).

Presence of a selectable marker gene in male sterility inducing constructs is also essential to facilitate *in vitro* selection of transformants and for field selection of plants containing the male sterility gene. For example, the *bar* gene from *Streptomyces hygroscopicus*, conferring resistance to the herbicide, Basta can be used as an *in vitro* as well as field selectable marker. Use of a strong constitutive promoter to express the marker gene is important for efficient selection of transformants. However, in a study on development of male sterile *barnase* lines in *Brassica juncea* (Indian mustard) (Jagannath *et al* 2001, Mol. Breeding 8:11-23), it was found that tissue-specific expression of the *barnase* gene was deregulated in the presence of a strong constitutive promoter (CaMV35S) that was used to express the herbicide resistance-conferring selectable marker gene (*bar*) in *barnase* constructs. Deregulated expression of the cytotoxic *barnase* gene not only reduced the recovery of transgenic shoots in transformation experiments, but also

affected agronomically important traits viz., vegetative morphology, female fertility, seed germination frequencies and inheritance of male sterility in *barnase* lines rendering them unsuitable for agronomic applications. To circumvent this problem, the above study described a method of using a Spacer DNA fragment to protect tissue-specific expression of the *barnase* gene which substantially enhanced the recovery of agronomically viable male sterile lines in *Brassica juncea*. Further, all male sterile lines developed in this study were found to be stable over several generations under field conditions in contrast to those developed earlier (Mariani *et al* 1990, Nature 347:737-741) in which several *barnase* lines showed breakdown of the male sterile phenotype even under controlled growth conditions (Denis *et al* 1993, Plant Physiol. 101:1295-1304).

Barnase activity in *B. amyloliquefaciens* is inhibited by a specific inhibitor, barstar, which negates the lethal effects of the ribonuclease by forming a one-to-one complex with the same in the cytoplasm (Hartley and Smeaton 1973, J. Biochem. 248:5624-5626). Expression of the *barstar* gene under the TA29 tapetum-specific promoter (which is also used for expression of the *barnase* gene to develop male sterile plants) was shown to restore fertility in *barnase* lines of *Brassica napus* (Mariani *et al* 1992, Nature 357:384-387). In the above study, four transgenic *barstar* lines containing a single copy of the T-DNA insert were crossed to four single-copy male sterile *barnase* lines in different combinations and resulting F1 hybrids were analyzed for restoration of male fertility in the same. It was found that six of nine crosses performed between *barnase* and *barstar* lines produced male fertile progeny while two crosses failed to restore fertility. In one case, partial restoration of fertility was also seen. Inability of some *barstar* lines to restore fertility was attributed to insufficient levels of the inhibitor protein due to its poor expression. However, the above study was found to be lacking in the analysis of pollen viability in restored plants. Hence, it is not possible to ascertain the extent of fertility restoration in male fertile progeny. This is of particular significance in case of oilseed crops such as *Brassicac*s wherein seed set is of paramount importance in achieving full yield potential of the crop.

Plant Genetic Systems have described, in EP 0412911A1, use of the *barstar* gene for restoration of fertility in male sterile *barnase* lines. The above patent describes development of restorer lines in tobacco and oilseed rape using a DNA construct containing three chimeric sequences (i) the *barstar* gene under transcriptional control of the TA29 tapetum-specific promoter (ii) a selectable marker gene (*neo*, conferring kanamycin resistance) expressed using a suitable constitutive promoter (pNos) for *in vitro*

selection of transformants and (iii) another selectable marker gene (*sf<sup>r</sup>*) expressed using the constitutive Rubisco ssu promoter from *Arabidopsis*. The above patent also describes various breeding strategies which can be used for hybrid seed production using the *barnase/barstar* system of male sterility and restoration. However, no experimental  
5 evidence has been provided for any of these strategies and therefore, the efficacy of the same cannot be conclusively established.

Another strategy for the generation of male sterile and restorer lines using the *barnase/barstar* system is described in US Patent 5,929,307 wherein the FLP/FRT recombinase system of yeast is used to regulate expression of the *barnase* and *barstar*  
10 genes. According to one of the embodiments of this invention, a male fertile transgenic plant can be generated by transformation using a vector containing a cytotoxic gene (*barnase*) and a restorer gene (*barstar*) wherein the restorer gene is flanked by site-specific (first) recombination sites which are recognized by a specific (first) recombinase. The male sterility and restorer genes are under independent transcriptional control of an  
15 anther-specific promoter. Further, the entire functional element composed of the above two genes and their respective promoters is flanked by another set of (second) site-specific recombination sites which are recognized by a specific (second) recombinase. A male sterile plant can be produced by crossing the above male fertile plant with a transgenic plant containing the first site-specific recombinase expressed under a constitutive  
20 promoter. Alternatively, the DNA construct used to generate the male fertile plant may also contain the first recombinase expressed under an inducible promoter, in which case, male sterility can be achieved by inducing expression of the first recombinase. Restoration of fertility can be achieved in either case by crossing the male sterile plant with a transgenic plant containing a gene encoding the second recombinase. The invention also  
25 describes several other alternative embodiments based on the use of site-specific recombinase systems to generate male sterile lines and for restoration of fertility in the same. However, data presented in the above study only establishes functionality of the yeast FLP/FRT system in maize cells and transformed *Arabidopsis thaliana* plants using various reporter genes. No experimental evidence is presented in support of functionality  
30 of the system in induction of male sterility and its restoration based on various strategies described in the same.

Novartis Finance Corporation has described, in US Patent 6,147,282, another strategy for controlling plant fertility based on chemical control of gene expression in plants. According to this strategy, chemical ligands can be used to activate receptor

polypeptides (for example, the Ecdysone and USP receptors containing ligand-binding and DNA-binding domains) which, in turn, induce expression of a target gene (*barnase* or *barstar*) expressed using a promoter containing suitable receptor binding elements. According to the patent, the above strategy could be used for induction of male sterility as well as its restoration. Male sterile transgenic plants can be generated using a DNA construct containing: (i) the *barnase* gene expressed under suitable anther-specific promoters (for example, TA29) (ii) the receptor expression cassettes (encoding the Ecdysone and USP receptor polypeptides) under regulatory control of the same anther specific promoter or from constitutive promoters (for example, maize ubiquitin, 35S or rice actin) and (iii) a target expression cassette having a 5' regulatory element (containing the appropriate response element sequence and core promoter elements) linked to the *barstar* gene. In order to restore fertility, a suitable chemical ligand (for example, RH5992) can be sprayed on the male sterile transgenic plants. This would induce heterodimerization of the USP and Ecdysone receptor polypeptides resulting in the activation of the 5' regulatory sequence of the target (*barstar*) expression cassette and production of barstar protein which would inhibit barnase and restore male fertility. The functionality of the above strategy was demonstrated using the *luciferase* reporter gene as the target polypeptide. Expression of luciferase protein was reported to increase approximately 2-4 folds upon induction over that observed in the absence of the chemical ligand. However, in the absence of corresponding data on expression of barstar protein for restoration of male fertility and in light of published literature on problems associated with development of male sterile lines due to deregulated expression of cytotoxic gene(s) (see above), efficiency of the above strategy in achieving the desired fertility status in crop plants needs to be established. Moreover, this strategy requires exogenous application of chemicals to induce the desired fertility status. For crop plants grown on a large acreage, this might not only be economically unsound, but the duration, time and amount of the desired ligand to be sprayed on standing crops would also be highly variable leading to inconsistency in production parameters.

Transgene expression levels can be enhanced by modification of the coding sequence to introduce preferentially used codons (for improved mRNA translation) and to ensure removal of potential deleterious signals which might inhibit transgene expression. Modified genes have been shown to improve transgene expression levels in heterologous systems including bacteria, plants and animal cell lines (Wosnick *et al* 1989, Gene 76:153-160; Adams *et al* 1988, Nucl. Acids Res. 16:4287-4298; Perlak *et al* 1991, PNAS-USA

88:3324-3328; Koziel *et al* 1993, BioTechnology 11:194-200; Pang *et al* 1996, Plant Physiol. 105:473-482). Common factors which have been taken into consideration for designing modified genes for improved expression in plant systems include codon usage, mRNA instability sequences (for example, polyadenylation signals), GC-content and plant  
5 intron consensus sequences. The *cry* genes of *Bacillus thuringiensis*, encoding the insecticidal  $\delta$ -endotoxins, are some of the best representative examples of bacterial genes modified for enhanced expression in transgenic plants (Perlak *et al* 1991, PNAS-USA 88:3324-3328; Adang *et al* 1993, Plant Mol. Biol. 21:1131-1145; Fujimoto *et al* 1993, BioTechnology 11:1151-1155; Strizhov *et al* 1996, PNAS-USA 93:15012-15017;  
10 Iannacone *et al* 1997, Plant Mol. Biol. 34:485-496). However, in a study by Rouwendal *et al* (1997) on the wild type and modified versions of the Green Fluorescent Protein (GFP) gene, no enhancement in expression levels could be achieved using modified transgenes in comparison to that obtained using the wild type sequence (Rouwendal *et al* 1997, Plant Mol. Biol. 33:989-999). In another study on the *luciferase* gene, it was  
15 reported that while the modified gene sequence may enhance expression of the transgene in one plant system, it may not do so in others (Lonsdale *et al* 1998, Plant Cell Rep. 17:396-399).

Plant Genetic Systems, in WO9810081, has described modification of the *barstar* gene sequence for enhanced expression in plant cells. According to the invention, the  
20 modified *barstar* DNA could be used for the development of efficient restorer lines which are capable of restoring fertility to a variety of male sterile lines (with varying levels of *barnase* expression) and are particularly useful for producing restorer lines in cereals, especially in corn, rice and wheat. The improved *barstar* gene sequence described in this invention encodes a barstar protein with an altered N-terminal amino acid sequence that  
25 starts with Methionine followed by "X" amino acid wherein "X" is either Alanine, Valine, Glycine, Aspartic acid or Glutamic acid, preferably Alanine encoded by a GCC codon. The preference for these amino acids is because, all available codons of these amino acids begin with a G nucleotide and being the second amino acid in the modified *barstar* gene sequence, it would provide an optimal translation initiation context at position +4. The  
30 modified *barstar* gene of the invention has an amino acid sequence which is "Met-Ala-Lys" while the wild type protein begins with "Met-Lys" at the N-terminal end. Other criteria used for designing the modified *barstar* gene sequence (also referred to as "synthetic *barstar* DNA") of the above invention and incorporated therein were:

- 5 (a) Codon usage : The modified sequence has a codon usage that is optimized for oilseed rape, cotton, maize, rice and wheat, preferably for oilseed rape (*Brassica napus*), maize and rice. For the purpose of this invention, the overall codon frequencies for various plant species described by Ikemura were used (Ikemura 1993, *In* "Plant Molecular Biology Labfax", Croy ed., Bios Scientific Publishers Ltd., pp. 37-48). According to this invention, the preferred codon selected for incorporation in the modified *barstar* sequence is one, which, in each of more than N/2 plant species (N=No. of plant species for which codon usage patterns are taken into account for design of the modified sequence) has an overall frequency that is at least twice the overall frequency of the least used codon and/or more than half of the overall frequency of the most used codon
- 10 (b) AT-content: A and T nucleotides constitute less than 40% of the sequence composition. This was primarily done to reduce the probability of introducing polyadenylation signals and intron recognition sequences which are known to destabilize transgene expression and can occur with a greater probability in AT-rich sequences. The synthetic *barstar* sequence of the invention has an AT-content of 38.4% while the wild type sequence is characterized by an AT-content of 51.6%.
- 15 (c) CG and CNG sequences: CG and CNG sequences have been shown to be associated with DNA methylation and gene silencing in plants. The modified sequence should not contain more than 7% CG dinucleotides and/or no more than 9.5% of CNG trinucleotides. The synthetic *barstar* DNA of the above invention contains 16 CG dinucleotides and 24 CNG trinucleotides while the wild type sequence has 14 CGs and 22 CNGs.
- 20 (d) The modified sequence is further characterized by having no more than 7, preferably no more than 5, particularly no more than 3 tetranucleotides consisting of only one kind of nucleotide and having no more than 2, preferably no more than 1 pentanucleotide consisting of only one kind of nucleotide. The synthetic *barstar* DNA of the invention contains one CCCCC and one GGGG stretch.
- 25 30

The improved *barstar* gene was deployed for the development of restorer lines in rice, corn and oilseed rape. In transgenic rice plants generated using the wild type and synthetic *barstar* genes under regulatory control of appropriate male tissue-specific promoters, it was found that the synthetic gene produced more barstar protein with the



amount of *barstar* protein being proportional to the amount of *barstar* mRNA present. Crosses between male sterile *barnase* lines and restorer *barstar* plants showed that restoration capabilities were directly comparable with the amount of *barstar* protein produced. Plants producing 50ng or more *barstar*/mg of extracted protein functioned as efficient restorer lines while plants producing between 4 – 10ng *barstar*/mg of extracted protein could only induce partial restoration of male fertility. While plants with *barstar* expression levels >50ng/mg of total protein could be obtained with the synthetic *barstar* gene, none of the plants generated using the wild type gene gave comparable expression levels. Similar studies were performed on transgenic corn plants. One male sterile line (designated MS3), which, according to the inventors, was a considerably difficult line to restore (data not shown), was selected for testing restoration capabilities of restorer lines generated using the wild type and synthetic *barstar* genes. Crosses between the male sterile MS3 line and seven wild type *barstar* restorer lines did not identify any suitable restorer, with maximally 75% of anthers producing viable pollen. On testing the MS3 line with restorer lines generated using the synthetic *barstar* sequence, it was found that four out of six lines tested were able to achieve complete restoration of male fertility thereby demonstrating that restorer lines generated using the synthetic *barstar* gene have a significantly better restoration capacity. The inventors further report that the amount of *barstar* protein in two restorer plants developed using the synthetic *barstar* DNA was much higher (210 and 100ng *barstar*/mg total protein) than that observed in one wild type *barstar* plant (20ng *barstar*/mg total protein). Transgenic plants were also generated in oilseed rape using the wild type and synthetic *barstar* genes. Analysis of activity of the wild type and synthetic *barstar* proteins revealed that activity of the improved *barstar* protein was at least equivalent to that of its wild type counterpart. However, no data were reported on restoration capabilities of these *barstar* lines. It is also important to note here that the above studies are lacking in data on copy numbers of the integrated transgene for the experimental lines used. Therefore, it is probable that position effects and varying copy numbers of the *barstar* gene may be influencing *barstar* expression levels in independent transgenic plants. Moreover, in the absence of data on pollen viability in restored plants, extent of restoration cannot be determined accurately.

Several strategies have been developed for generation of restorer lines for hybrid seed production. One of these studies has also addressed the issue of expression level of the restorer gene product and has prescribed the design and use of synthetic DNA sequences for enhanced expression in heterologous systems to facilitate effective

restoration. However, none of the described strategies highlight the extent of restoration with reference to pollen viability which is a critical parameter for determining the efficacy of any restoration strategy. Additionally, most strategies rely on the use of several superfluous genes, promoters and other DNA sequences, the presence of which is highly undesirable in transgenic crop systems.

The present invention describes a method for enhancing expression levels of transgene(s) while reducing its susceptibility to homology based post transcriptional gene silencing. It further describes a method for achieving stable, enhanced and extended temporal expression of a restorer gene product based on the simultaneous use of two different restorer gene sequences encoding the same protein product in the same DNA construct. One of the said sequences is modified for expression in dicotyledonous crop plants using codon degeneracy to avoid homology between the two sequences at the DNA and mRNA levels and the other is the naturally occurring wild type sequence, each of which is placed under independent transcriptional control of regulatory elements with overlapping expression patterns in male reproductive tissues. In addition to enhancing restorer gene expression, simultaneous use of two different DNA sequences (which would transcribe two different RNA molecules) in the same DNA construct confers the added advantage of avoiding homology based post transcriptional gene silencing. The method described in the present invention can also be applied with suitable modifications to other molecular methods based on the expression of a cytotoxic gene and its corresponding inhibitor for the development of male sterile and restorer lines, respectively.

## OBJECTS OF THE INVENTION

The main object of the invention is to provide a novel method for enhancing expression levels of transgene(s) with reduced susceptibility to homology based post transcriptional gene silencing by the simultaneous use of two different DNA sequences encoding the same protein product.

Another object of the invention is to provide a method for achieving stable, enhanced and extended temporal expression of a restorer gene product in male reproductive tissues for the development of improved restorer lines in crop plants for hybrid seed production.

Yet another object is to provide a modified sequence of a restorer gene designed for expression in dicotyledonous crop plants.

A further object is to provide a DNA construct containing two different coding sequences of a restorer gene - one modified for expression in dicotyledonous crop plants using codon degeneracy to avoid homology between the two sequences at the DNA and mRNA levels and another which is the naturally occurring wild type sequence – each  
5 under independent transcriptional control of regulatory elements with overlapping expression patterns in male reproductive tissues, the protein product generated using the wild type and modified sequences remaining identical.

Another object is to develop a DNA construct that would enhance expression of the restorer gene product while providing protection against post-transcriptional gene  
10 silencing.

Yet another object is to develop improved restorer lines using the DNA construct of the invention.

Still another object is to develop methods for the identification of improved restorer lines generated using the DNA construct of the invention.

15

## SUMMARY OF INVENTION

The present invention relates to a method for enhancing expression levels of transgene(s) while reducing its susceptibility to homology based post transcriptional gene silencing. The invention provides a novel method for achieving stable, enhanced and  
20 extended temporal expression of a restorer gene product in male reproductive tissues for the development of improved fertility restorer lines in crop plants for hybrid seed production. The invention also provides a modified restorer gene sequence designed for expression in dicotyledonous crop plants using codon degeneracy to avoid homology between the modified sequence and the wild type sequence at the DNA and mRNA levels.

25 Further, the invention provides a DNA construct comprising: (i) a first transcription unit containing the wild type sequence of a restorer gene under transcriptional control of a suitable first tapetum-specific promoter (which is used to express the male sterility gene in corresponding male sterile plants) and fused to a suitable transcription termination signal including a polyadenylation signal, (ii) a second transcription unit comprising a modified  
30 sequence of the restorer gene, designed for expression in dicotyledonous crop plants, under transcriptional control of a second tapetum-specific promoter and fused to a suitable transcription termination signal including a polyadenylation signal and (iii) a third transcription unit comprising a selectable marker gene under transcriptional control of a strong constitutive promoter and fused to a suitable transcription termination signal

including a polyadenylation signal. The second tapetum-specific promoter in the DNA construct of the invention is selected such that the first and second tapetum-specific promoters have overlapping expression profiles. Further, the invention describes methods for the development and identification of improved restorer lines using the said DNA  
5 construct.

## DETAILED DESCRIPTION OF INVENTION

The present invention relates to a method for enhancing expression levels of transgene(s) while reducing its susceptibility to homology based post transcriptional gene  
10 silencing. It further describes a method for achieving stable, enhanced and extended temporal expression of a restorer gene product in male reproductive tissues and a method for the development of improved fertility restorer lines in crop plants for hybrid seed production. The method of the present invention is based on the simultaneous use of two different restorer gene sequences – one being the naturally occurring wild type sequence  
15 and the other sequence generated by modification of the wild type sequence for expression in dicotyledonous crop plants by using codon degeneracy to avoid homology between the modified sequence and the wild type sequence at the DNA and mRNA levels – each of the said sequences being placed under independent transcriptional control of different overlapping plant tissue-specific regulatory elements in the same DNA construct.  
20 Accordingly, the invention provides a DNA construct comprising:

- (i) a first transcription unit comprising the wild type sequence of a restorer gene under transcriptional control of a suitable first tissue-specific promoter (which is used to express the male sterility gene in corresponding male sterile plants) and fused to a suitable transcription termination signal, including a polyadenylation  
25 signal,
- (ii) a second transcription unit comprising a modified sequence of the restorer gene, designed in the present study for expression in dicotyledonous crop plants, under transcriptional control of a second tissue-specific promoter and fused to a suitable transcription termination signal including a polyadenylation signal, and  
30
- (iii) a third transcription unit comprising a selectable marker gene under transcriptional control of a strong constitutive promoter and fused to a suitable transcription termination signal including a polyadenylation signal.

In an embodiment, the restorer genes that can be used in the DNA construct of the invention could be any coding sequence which can inhibit or negate the cytotoxic effects of another DNA sequence or its encoded product thereby protecting the cell from lethal effects of the cytotoxic gene. Instances of such restorer genes include *barstar* (inhibitor of the ribonuclease, barnase) or protease inhibitors (which neutralize the activity of proteases such as papain). In a preferred embodiment, the restorer gene used in the present invention is the *barstar* gene from *Bacillus amyloliquefaciens* which functions as an intracellular inhibitor of barnase. In a preferred embodiment of the invention, two different versions of the *barstar* gene are used in independent transcription units for stable, enhanced and extended temporal expression of the restorer gene product. One of the sequences is represented by the wild-type gene and the other is a modified sequence designed for expression in dicotyledonous crop plants.

The modified sequence of the restorer gene (*barstar*) was designed based on parameters known to influence transgene expression in heterologous systems. Critical parameters taken into consideration for designing the modified sequence and changes incorporated therein are described hereunder.

(1) Codon usage:

Codon usage patterns have been well characterized for several organisms viz. *E.coli*, *Bacillus*, *Agrobacterium*, yeast, *Drosophila*, man and higher plants. Codon usage variations have a significant impact on gene expression in heterologous systems. Hence, while designing genes for enhanced expression, it is important to replace codons that are least favored in the target system with appropriate synonymous codons to facilitate improved translatability of mRNA. Based on published data on codon usage patterns for various organisms (Murray *et al* 1989, Nucl. Acids Res. 17:477-498; Wada *et al* 1992, Nucl. Acids Res. 20 Supplement:2111-2118; Nakamura *et al* 1997, Nucl. Acids Res. 25:244-245), a codon usage table for dicotyledonous plants was developed to design a modified sequence of the *barstar* gene (Table 1). For clarity and convenience in selection of codons for incorporation in the modified sequence, all codons for each amino acid were categorized (wherever possible) into 3 classes:

1. Optimal codons : that have maximum usage in the system.
2. Sub-optimal codons: with an intermediate usage between the Optimal and Rare codons
3. Rare codons : with a usage less than 50% of the usage of the Optimal codons.

Table 1. Codon usage table for dicot plants

Amino acid	Optimal codon(s)	Sub-optimal codon(s)	Rare codon(s)
Arginine	AGA AGG	CGT	CGC CGA CGG
Leucine	CTT TTG	CTC	TTA CTA CTG
Serine	TCT	TCC TCA AGC AGT	TCG
Threonine	ACT	ACC ACA	ACG
Proline	CCT CCA	—	CCC CCG
Alanine	GCT	GCA GCC	GCG
Glycine	GGA GGT	—	GGG GGC
Valine	GTT	GTG	GTC GTA
Lysine	AAG	—	AAA
Asparagine	AAC	AAT	—
Glutamine	CAA	CAG	—
Histidine	CAT	CAC	—
Glutamic acid	GAG GAA	—	—
Aspartic acid	GAT GAC	—	—
Tyrosine	TAC TAT	—	—

Cystine	TGT TGC	—	—
Phenyl alanine	TTC TTT	—	—
Isoleucine	ATT	ATC	ATA
Methionine	ATG	n.a.	n.a.
Tryptophan	TGG	n.a.	n.a.

n.a. : not applicable

In an embodiment of the present invention, forty seven codons (of a total of 91 codons) from the wild type *barstar* gene were replaced by preferentially used codons in the modified sequence. On the basis of the three-tier classification of codons for dicot systems described above, the wild type gene was found to have a codon usage comprising 50 optimal codons, 25 sub-optimal codons and 15 rare codons (excluding the termination codon). The modified sequence on the other hand, is composed of 78 optimal codons and 12 sub-optimal codons with complete exclusion of rare codons. Retention of 12 sub-optimal codons was primarily due to consideration of other parameters (as described below) for designing a modified gene sequence.

## (2) Potential mRNA instability sequences and plant poly(A) signals:

An important mechanism regulating gene expression operates at the level of mRNA turnover. Several sequences governing instability of mRNAs have been studied and characterized. For example: the 40bp DST sequences found downstream of the coding region in Small Auxin Up RNA (SAUR) transcripts of several plants (McClure *et al* 1989, Plant Cell 1:229-239) and the AUUUA repeat motif in the 3' untranslated regions (UTR) of several mammalian (Peltz *et al* 1991, *In Critical Review, Eukaryotic Gene Expression*, Eds. Stein JS, Stein JL and Lian JB, CRC Press) and plant mRNAs. AT-richness has also been implicated in modulating mRNA stability in plants. However it has been shown that the effect of AT-rich sequences on mRNA stability is sequence-specific (consisting of ATTTA repeats) and not simply due to high AT-content (Ohme-Takagi *et al* 1993, PNAS-USA 90:11811-11815).

Sequences favoring mRNA stability have not yet been well characterized with the exception of the poly(A) tail. Poly(A)-binding proteins (PABPs) have been implicated in governing mRNA stability. Available data suggest that plant PABPs may

play an important role in organ-specific, post-transcriptional control of gene expression by influencing mRNA stability. The most widely used Plant Polyadenylation Signal Sequence (PPSS) is AATAAT although several others, a representation of which is given below, have also been proposed to be functional at low frequencies (Dean *et al*  
 5 1986, Nucl. Acids Res. 14:2229-2240):

$$A_{15} \ A_9 \ T_6 \ A_8 \ A_{13} \ T_8$$

$$T_6 \ A_5 \ C_6 \ T_2 \ A_7$$

10  $G_3 \ T_1$

$$C_1$$

These consensus sequences were derived from a study of 15 sequences  
 15 representing cDNA clones of *rbcS* and *cab* genes from petunia and *bronze* genes from maize. Numbers in subscripts indicate the relative preference for a nucleotide at a particular position in the poly(A) signal sequence of the 15 sequences analyzed.

Presence of such signal sequences in the coding sequence of a gene might lead to premature termination of mRNA synthesis leading to unstable expression of the target  
 20 gene. Therefore, it is advisable to remove such putative mRNA instability sequences as well as PPSS sequences from the coding regions of transgenes to prevent any adverse effects on transcript stability/expression. Consensus sequences for the above types of signals were absent in the wild type sequence and were also not created during design of the modified gene sequence.

25

### (3) Plant intron consensus sequences:

Plant introns are characterized by the presence of 5' and 3' splice sites and a branch point, all of which have well-characterized consensus sequences (Wiebauer *et al*  
 1988, Mol. Cell. Biol. 8:2042-2051; Goodall and Filipowicz 1989, Cell 58:473-483;  
 30 Schuler 1998, *In A Look Beyond Transcription: Mechanisms Determining mRNA Stability and Translation in Plants*, Bailey-Serres J and Gallie DR, Eds. American Society of Plant Physiologists; see Table below).



Motif	Sequence
5' Splice site	AG / <u>G</u> TAAAGT
Branch point	CTN <u>A</u> N TTN <u>A</u> N
3' Splice site	GCAG / G

Underlined nucleotides represent bases that are highly conserved in all plant introns. Abbreviations: A: Adenine, T: Thymine, G: Guanine, C: Cytosine, N: A,T,G or C, Y: Cytosine or Thymine, R: Adenine or Guanine.

5 In order to prevent any undesirable post-transcriptional processing of the transgene mRNA, it is necessary to remove potential 5'/3' splice sites and branch points from the coding region of a transgene. The wild type sequence of the *barstar* gene carries three branch point consensus sequences and one 3' splice site consensus sequence which were eliminated from the modified sequence during gene design.

10

#### (4) Methylation-prone sequences:

Transgene silencing in plants has almost always been found to be associated with methylation of the transgene sequences at 'CG' and 'CNG' residues and involves transcriptional inactivation (which is meiotically heritable) as well as post-transcriptional processes (which are reset after meiosis). Methylation of both upstream regulatory sequences (Matzke *et al* 1989, EMBO J. 8:643-649; Linn *et al* 1990, Mol. Gen. Genet. 222:329-336; Matzke and Matzke 1991, Plant Mol. Biol. 16:821-830; Kilby *et al* 1992, Plant Mol. Biol. 20:103-112; Meyer *et al* 1993, Plant J. 4:89-100; Matzke *et al* 1994, Mol. Gen. Genet. 244:219-229; Meyer and Heidmann 1994, Mol. Gen. Genet. 243:390-399) as well as coding regions of transgenes (Meyer and Heidmann 1994, Mol. Gen. Genet. 243:390-399; Meyer *et al* 1993, Plant J. 4:89-100) has been reported in several transgenic plant systems. To minimize the influence of 'CG' and 'CNG' residues in mediating transgene silencing, it is advisable to remove them from the coding regions of transgenes by using codon degeneracy. There are, however, several codons which, although inherently lack a 'CG' or a 'CNG', can introduce such sites when the amino acids encoded by such codons occur in tandem in the protein sequence. For example: ala-ala, ala-aspartic, pro-ala, ala-gly, ala-glu, thr-gly, pro-gly, ala-val, pro-glu, thr-glu, thr-val, pro-aspartic, thr-aspartic and pro-val. In such situations, it is not possible to remove CG/CNG sites using codon degeneracy. The number of CGs and CNGs has been reduced in the modified *barstar* gene

15

20

25

sequence to 3 and 8 respectively from the corresponding figures of 14 and 22 in the wild type sequence. Complete removal of CGs and CNGs from the modified sequence has not been possible due to the tandem arrangement of some amino acids as described above.

5 (5) GC-content:

Analysis of gene sequences from plant systems has revealed that plant genes have a characteristic GC-content which is seen to vary between monocot and dicot species. While monocot genes have an average GC-content of 60-70%, dicot genes are characterized by a GC-content of 42-46% (Salinas *et al* 1988, Nucl. Acids Res. 16:4269-  
10 4285). GC-content also plays an important role in influencing DNA methylation. Genes located in a GC-rich region are characterized by relatively higher levels of methylation as compared to those in AT-rich regions. Significant variations in GC-content between the transgene and the endogenous genes at the site of integration may render the former susceptible to methylation-induced silencing. Hence, while designing  
15 genes for enhanced expression in a heterologous system, it is advisable to maintain GC-content of the transgene around the average prescribed GC-content of the expression system.

The wild-type transgene sequence was also analyzed for the presence of AT- and GC-rich sequences and those greater than 4 nucleotides in length were removed from the  
20 modified sequence (without compromising on other critical parameters) during gene design. This parameter assumes significance because long stretches of AT/GC might not only interfere with design of suitable oligos for gene synthesis but also influence mRNA processing (stability) and GC-content of the transgene. Reduction in the number of such sequences would serve to ensure a more equitable distribution of various nucleotides over  
25 the gene sequence rendering it more homogeneous in its GC-content.

Following incorporation of all the aforementioned modifications, the GC-content of the modified sequence was 42.1% while that of the wild-type sequence was 48.4%. Both the wild type and modified genes have a GC-content that falls within the average range of %-GC for dicot genes (42-46%).

30

(6) Consensus translational start site

The context of the ATG codon of a gene is known to play a significant role in influencing translation initiation (Kozak 1986, Cell 44:283-292). A consensus sequence for this region has been arrived at on the basis of an extensive study of several plant genes

by Kozak (Kozak 1984, Nucl. Acids Res. 12:857-872) and Joshi (Joshi 1987, Nucl. Acids Res. 15:6643-6653). The consensus sequence reported by Kozak (1984) has been incorporated (with a minor change) in the context of the first ATG codon of the modified *barstar* gene as shown below.

- 5 Context of ATG codon of modified *barstar* gene : 5' – CCACC ATG AAG – 3'  
Met Lys  
Kozak's consensus for plant genes : 5' – CCACC ATG G – 3'  
Met

10 This change became necessary in view of the non-availability of a codon beginning with 'G' for Lysine.

Alterations effected in the modified *barstar* gene sequence are summarized in Table 2 below.

Table 2: Summary of modifications made in the modified *barstar* gene sequence vis-à-vis the wild-type sequence.

S.No.	Parameter	Wild-type <i>barstar</i> gene	Modified <i>barstar</i> gene
1.	Potential poly(A) signals	0	0
2.	mRNA instability sequence [(ATTTA) <sub>n</sub> ]	0	0
3.	Methylation-prone sequences		
	CG	14	3
	CNG	22	8
4.	Plant intron consensus sequences (*)		
	a) Branch point sequences		
	YNYRAY	1	0
	CYRAY	2	0
	b) 5' splice site (AAGGTAAGT)	0	0
	c) 3' splice site (TNYAGG)	1	0
5.	AT-stretches (WWWW)	17	5
6.	GC-stretches (SSSS)	12	0
7.	GC-content	48.4%	42.1%

- 15 \* : consensus sequences available till 1996 have been taken into consideration for gene design in the present study; W : Adenine or Thymine; S : Guanine or Cytosine.

In an embodiment of the invention, the wild type and modified sequences of the *barstar* gene are expressed in a tissue-specific manner using tapetum-specific promoters which have overlapping expression profiles in the target tissues. Tissue-specific promoters are characterized by definite temporal and spatial expression patterns during plant growth and development. Several such promoters have been characterized from plant systems. A few examples include anther-specific promoters such as TA29, A9, *tap1*, *bcp1* or seed specific promoters such as *napin* or promoters such as AP3 which express in a relatively wider manner in floral tissues. The tissue-specific promoters in the DNA construct of the present invention are selected such that the first and second tissue-specific promoters have overlapping expression profiles. In a preferred embodiment, the first tissue-specific promoter used to express the wild type *barstar* sequence is the tapetum-specific promoter, TA29 (Seurinck *et al* 1990, Nucl. Acids Res. 18:3403), while the second tissue-specific promoter used to express the modified sequence of *barstar* gene is another tapetum-specific promoter, A9 (Paul *et al* 1992, Plant Mol. Biol. 19:611-622).

A selectable marker gene encodes an RNA or protein which, when expressed in the cells of a plant, gives the cells expressing the gene a selective advantage over cells lacking the same. In an embodiment, the marker gene is selected from the group of herbicide resistance-conferring genes such as *bar*, *tfdA*, *ALS*; antibiotic resistance-conferring genes such as *npIII*, *hpt*, *aadA*, etc. In a preferred embodiment of the invention, the selectable marker gene used for *in vitro* and *in vivo* selection of transformed tissues is the herbicide resistance conferring gene, *bar*, from *Streptomyces hygroscopicus*. In another preferred embodiment of the invention, the marker gene used for *in vitro* selection is the *hpt* gene conferring resistance to hygromycin.

In an embodiment, the promoter used for driving the expression of the *bar* gene is the CaMV35S promoter with a duplicated enhancer (henceforth referred to as CaMV35Sde) which is a known strong constitutive promoter. The duplication of the enhancer region (-343 to -90) has been shown to enhance expression of the CaMV 35S promoter by 10-fold (Kay *et al* 1987, Science 236:1299-1304). In another embodiment, the promoter used for expression of the *hpt* gene is the CaMV35S promoter with a single enhancer.

The transcription unit of the selectable marker gene (*bar* or *hpt*) is placed towards the Left Border of T-DNA (in an appropriate binary vector) to ensure complete transfer of all components located between the T-DNA borders during *Agrobacterium*-mediated genetic transformation.

The method outlined in the present invention for enhanced expression of a transgene with reduced susceptibility to homology based gene silencing is based on the simultaneous use of two different sequence versions (wild type and modified) of a restorer gene, both of which encode identical protein products and are placed under independent transcriptional control of two different regulatory elements having overlapping expression profiles. The efficacy of the above method in developing efficient restorer lines by enhancing expression of the restorer gene product can be tested either by crossing transgenic *barstar* lines generated using constructs described in the present invention with male sterile *barnase* lines or by retransforming *barnase*-containing male sterile plants with *barstar* constructs. In the latter approach, to facilitate efficient selection of transformants containing both the male sterility and restorer genes, the marker gene used in the DNA construct of the present invention could be different from that present in the male sterile transgenic plant(s). For example, male sterile plants containing the *bar* gene as a selection marker can be transformed with restorer DNA constructs containing the *hpt* gene for selection of transformants. A similar approach can be used for other combinations of marker genes used for generation of male sterile and restorer plants. The frequency of male fertile transformants among transgenic plants containing both the marker genes (i.e., both the *barnase* and *barstar* genes) would reflect the efficacy of the strategy in development of efficient restorers for male sterile lines in crop plants.

The efficacy of the aforementioned strategy was tested by transforming a stable, male sterile *barnase* line (*bn* 3.6; containing the *bar* gene as a selectable marker; Jagannath *et al* 2001, Mol. Breeding 8:11-23) with a DNA construct containing the wild type and modified sequences of the *barstar* gene under transcriptional control of the TA29 and A9 promoters respectively and the *hpt* gene under transcriptional control of the CaMV35S single enhancer promoter as a selectable marker. As appropriate controls, the male sterile plant was also transformed with DNA constructs containing a single copy of the wild type *barstar* gene under transcriptional control of either the TA29 or A9 promoter. Transformants containing both the *barnase* and *barstar* genes were obtained by sequential selection on selective media containing hygromycin (from *barstar* T-DNA) and Basta (Agrevo; from *barnase* T-DNA). Transgenic plants were transplanted in a nethouse under field conditions during the growing season (October-April) and analysed for their male fertility/sterility phenotype.

Use of the *barstar* gene as a restorer of male fertility requires that it is expressed equally or to a greater extent than the *barnase* gene. Transgenic plants wherein *barstar*

expression levels are sufficient to counter the lethal effects of barnase would be male fertile. Since barnase levels are identical in all transformation experiments with both control and test DNA constructs, any variation in the frequency of fertile (i.e., restored) plants would be primarily due to differences in barstar levels in the same.

- 5 Restoration of male fertility in transformed plants containing both the *barnase* and *barstar* genes was analyzed on the basis of pollen formation, pollen viability and formation of selfed seeds and correlated with copy number of the *barstar* gene cassette in the same. Results on the frequency of fertile (restored) plants obtained from transformation of the *barnase* line with various *barstar* constructs are summarized in
- 10 Table 3 below.

Table 3 : Frequency of fertility restoration in *bn* 3.6 using *barstar* constructs

Construct Code	Construct	No. of Basta <sup>R</sup> - Hyg <sup>R</sup> plants analysed in field	Fertility status			Percent of restored plants
			F	S	SF	
A	A9- <i>bs(fm)</i> :: TA29- <i>bs(wt)</i>	88	79	8	1	89.8
B	TA29- <i>bs(wt)</i>	64	42	15	7	65.6
C	A9- <i>bs(wt)</i>	47	6	40	1	12.7
D	A9- <i>bs(fm)</i>	46	5	38	4	10.8

F: Fertile ; S : Sterile ; SF : Semi-fertile ; *bs(wt)* : wild type *barstar* gene ; *bs(mod)* : modified *barstar* gene

- 15 The above results indicate that restoration of fertility using constructs containing a single copy of the wild type or modified *barstar* gene (Constructs C, D) under transcriptional control of the A9 promoter was very low (12.7% and 10.8%; Table 3). The frequency of obtaining fertile plants with the TA29-*bs(wt)* construct (Construct B) was much higher (65.6%). This difference in restoration capability can be attributed to the fact
- 20 that the *barnase* gene in the male sterile parent is under transcriptional control of the TA29 promoter. Therefore, in such plants transformed with the TA29-*bs(wt)* construct, both the *barnase* and *barstar* genes would have a similar temporal expression profile leading to simultaneous production of both the barnase and barstar proteins. Expression from the A9 promoter, on the other hand, ends earlier than that of the TA29 promoter. Therefore, no
- 25 barstar protein would be available to inhibit the lethal effects of barnase during later stages of TA29-*barnase* expression and plants would therefore be male sterile.

On using construct A [*A9-bs(fm)::TA29-bs(wt)*], the frequency of fertile plants registered a substantial increase to 89.8% indicating that barstar levels are significantly higher as compared to the earlier (control) constructs. In fertile plants derived using construct A, the presence of two independent transcriptional units for the *barstar* gene leads to the formation of more *barstar* mRNA and hence more barstar protein. Since the tapetum-specific promoters used to transcribe the *barstar* gene have overlapping expression profiles, accumulation of barstar protein in tapetal tissues begins earlier (from the A9 promoter) than barnase (which is under transcriptional control of the TA29 promoter) and continues during the entire period when barnase is expressed (from the TA29-*barstar* cassette). Extended expression of the *barstar* gene therefore builds up a reservoir of the inhibitor protein which ensures effective inhibition of barnase.

The occurrence of a few fertile plants among transformants generated using Constructs C and D could be due to position effect-mediated enhancement of *barstar* gene expression and these were not taken further for subsequent analysis. Similarly, the occurrence of semi-fertile and male sterile plants could be attributed to position effect-mediated inhibition of barstar expression. It is nevertheless also important to note that the occurrence of semi fertile plants among transformants generated using Construct A is significantly lower than those obtained with Construct B indicating the greater efficiency of Construct A in generation of male fertile plants.

To determine the extent of fertility restoration, male fertile (restored) plants obtained using Constructs A and B were analyzed for their pollen viability. Fertile, restored plants derived from each of the above constructs were classified into the following three categories on the basis of their pollen viability: < 50%, 50 – 90%, >90%. In case of efficient restoration, pollen viability in restored events is expected in the range of 90 – 100% similar to that observed in untransformed control plants or in transgenic plants lacking the *barnase* gene. Pollen viability data of transformed male fertile (restored) plants is summarized in Table 4 below.

Table 4 : Pollen viability of male fertile, restored events

Construct code	Construct	No. of fertile plants analysed	No. of plants with pollen viability			% of plants with PV > 90%
			<50%	50 - 90%	>90%	
A	<i>A9-bs(fm) :: TA29-bs(wt)</i>	79	3	24	52	65%
B	<i>TA29 :: bs(wt)</i>	42	1	21	20	47.6%

PV: pollen viability

From the above data, it is clear that fertile (restored) plants with pollen viability >90% were obtained at a greater frequency using Construct A (65%) as compared to those obtained using Construct B (47.6%; Table 4).

5 To determine if better restoration capabilities using Construct A were due to higher copy number, fertile (restored) plants were subjected to Southern analysis. The data indicated that there was no bias towards multiple copy integrations among fertile (restored) plants generated using any of the three constructs thereby indicating that the frequency of fertile (restored) plants was a direct reflection of construct efficacy alone.  
10 Thus, among the constructs studied, Construct A, containing two different versions of the restorer gene, was found to function more effectively than Construct B (containing a single copy of the wild type gene sequence). All fertile plants thus obtained also showed normal seed set on selfing.

To further ascertain the efficacy of the strategy outlined in the present invention,  
15 transgenic *barstar* plants were generated using a DNA construct (hereafter referred to as Construct I) containing the wild type and modified sequences of the *barstar* gene under transcriptional control of the TA29 and A9 promoters respectively and the *bar* gene driven by the CaMV35S double enhancer promoter as a selectable marker. As an appropriate control, transgenic plants were also generated using a DNA construct containing a single  
20 copy of the wild type *barstar* gene under transcriptional control of the TA29 promoter (Construct II). Transformants were selected on media containing Basta (Agrevo). Transgenic plants generated using the above constructs were subjected to Southern analysis to identify single copy events which were subsequently crossed to 4 single copy male sterile *barnase* lines. F1 progeny derived from the above crosses (40-60 plants/cross)  
25 were analyzed for their male fertility/sterility status to determine the restoration abilities of various *barstar* lines. Results of this study are summarized in Table 5 below. Numbers in each cell represent the number of restorer lines obtained of the total number of *barstar* lines tested for a given *barnase* line.

30



Table 5. Restoration profile of *barstar* transgenic plants

Construct Code	<i>Barstar</i> constructs	<i>barnase</i> lines				No. of restorers identified/ No. of lines tested
		<i>bn</i> 3.6	<i>bn</i> 3.23	<i>bn</i> 3.4	<i>bn</i> 3.48	
I	A9- <i>bs(fm)</i> :: TA29- <i>bs(wt)</i>	8/11	6/11	7/11	4/11	25/44
II	TA29 :: <i>bs(wt)</i>	2/4	0/4	0/4	0/4	2/16

The above observations conclusively establish that Construct I, containing two different versions of the restorer gene sequence, provides restorer lines at a much greater frequency (25 restorers identified out of 44 lines tested; ~57%) than Construct II (2 of 16 lines tested; ~12%). Furthermore, a greater number of restorer lines could be identified for each *barnase* line among *barstar* plants generated using Construct I than those obtained with Construct II. Of the four *barnase* lines tested with Construct II, restorers could be identified only for *bn* 3.6.

Thus, the above results clearly indicate that the method of using two different versions (wild type and modified) of a gene sequence (encoding identical protein products) in the same DNA construct effectively enhances transgene expression levels. It is emphasized here that modifications of the gene sequence as described in this invention can be made for any other gene sequence for both constitutive as well as tissue-specific enhancement in expression using appropriate regulatory elements. Parameters for such modifications can be suitably altered according to the target system in which enhanced expression is desired. The design of the modified *barstar* gene sequence, its simultaneous use with the corresponding wild type sequence in a DNA construct for enhancing expression of the restorer gene product and the use of the TA29 and A9 promoters as overlapping regulatory elements have been cited merely as a representative example and should not be construed to limit the scope of this invention.

The invention also provides a method to obtain efficient fertility restorer lines for male sterile plants for hybrid seed production. The said method for the development of efficient fertility restorer lines comprises the steps of:

- i) transforming the nuclear genome of plant cells with a DNA construct (hereafter referred to as Construct I) comprising:

a) a first transcription unit comprising the wild type sequence of a restorer gene under transcriptional control of a suitable first tissue-specific promoter (which is used to express the male sterility gene in corresponding male sterile plants) and fused to a suitable transcription termination signal, including a polyadenylation signal.

b) a second transcription unit comprising a modified sequence of the restorer gene, designed in the present study for expression in dicotyledonous crop plants, under transcriptional control of a second tissue-specific promoter and fused to a suitable transcription termination signal including a polyadenylation signal, and

c) a third transcription unit comprising a selectable marker gene under transcriptional control of a strong constitutive promoter and fused to a suitable transcription termination signal including a polyadenylation signal.

ii) regenerating plants from said transformed plant cells,

iii) identifying single copy plants by Southern hybridization,

iv) crossing the above single copy plants with male sterile *bar* lines,

v) analyzing segregating F1 progeny to identify plants containing the marker gene,

vi) analyzing segregation of male-fertile and male-sterile phenotypes among marker gene-containing F1 progeny to identify putative male sterile-restorer combinations,

vii) molecular analysis of F1 progeny by Tissue-PCR to identify fertility restored plants,

viii) testing pollen viability of fertility restored plants to determine extent of restoration,

ix) selfing the restored plants to obtain F2 progeny

x) analyzing F2 progeny under field conditions for segregation of male-fertile and male-sterile phenotypes to confirm male sterile-restorer combination.

In an embodiment, the preferred restorer gene product is *barstar*.

In still another embodiment, the preferred first tissue-specific promoter is TA29.

In yet another embodiment, the preferred second tissue-specific promoter is A9.

In another embodiment, the preferred marker gene is *bar*.

In an embodiment, the preferred constitutive promoter is the CaMV35S promoter with a duplicated enhancer.

In an embodiment, the crop plants used for genetic transformation are selected from dicotyledonous plants, such as *Brassica juncea*.

In another embodiment, the procedure used for development of transformed plants in *Brassica juncea* is *Agrobacterium*-mediated transformation using disarmed Ti plasmid.

In an embodiment, the transgenic plants are analyzed by Southern hybridization to identify transgenic plants containing a single copy of the T-DNA insert.

5 In yet another embodiment, single copy plants containing the wild type and modified sequences of the restorer gene are crossed with single copy male sterile *barnase* lines to analyze their restoration capability.

10 In another embodiment, the F1 progeny obtained from selected backcrossed progeny were transferred to field conditions and tested for stable segregation of male sterility and fertility among maker gene-containing plants.

In an embodiment, fertility restored events among segregating F1 progeny are identified by Tissue PCR.

In another embodiment, fertility restored events were analyzed for pollen viability and selfed to obtain F2 progeny.

15 In yet another embodiment, F2 progeny grown under field conditions were analyzed for segregation of male-fertility and male-sterility to confirm the male sterile-restorer combination.

20 The invention is described in detail hereinafter, with reference to the accompanying drawings and examples. Modifications, especially with respect to the DNA construct, its various components viz., the restorer gene(s), tissue-specific promoters, marker genes, constitutive promoters used to express the marker gene and the methods used to deploy the same would be obvious to those skilled in the art. Such modifications are deemed to fall within the scope of this invention and the examples and embodiments provided herein should not be construed as limitations on the inventive concept embodied  
25 in this invention.

## BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWINGS

In the drawings that accompany the specification

30 Fig. 1: is a representation of the number of optimal, sub-optimal and rare codons among the wild type and modified *barstar* gene sequences. The X-axis represents distribution of codons among the wild-type and modified genes. All codons for each amino acid have been categorized (wherever possible) into three classes in accordance with their usage status in plant systems:

(1) Optimal codons : with maximum usage

- (2) Sub-optimal codons : with an intermediate usage  
 (3) Rare codons : with a usage less than 50% of the usage of optimal codons

The Y-axis represents the total number of codons in each category. The wild-type *barstar* gene has 50 optimal codons, 25 sub-optimal codons and 15 rare codons (excluding the termination codon) while the modified sequence has a codon usage consisting of 78 optimal codons and 12 sub-optimal codons. Rare codons have been completely excluded from the modified sequence.

Fig. 2A : is the naturally occurring wild type nucleotide sequence of the *barstar* gene. (Seq. ID # 1). Putative plant intron consensus sequences in the wt gene are highlighted in grey boxes (NNNNNNN).

Fig. 2B: is the amino acid sequence of the protein of the naturally occurring wild type *barstar* gene. (Seq. ID # 2)

Fig. 3A : is the modified nucleotide sequence of the *barstar* gene. (Seq. ID # 3)

Fig 3B is amino acid sequence of the *barstar* gene whose nucleotide sequence has been modified. (Seq. ID # 4). It will be noted that this sequence is same as Seq. ID # 2

Fig. 4 : is a comparative representation of the wild type (wt) and modified (mod) sequences of the *barstar* gene. Putative plant intron consensus sequences in the wt gene are highlighted in grey boxes (NNNNNNN).

Fig. 5 : is a schematic representation of a DNA Construct (Construct I) of the invention depicting various components and locations of restriction enzyme sites. Abbreviations: LB – Left Border of T- DNA of *Agrobacterium tumefaciens*, RB – Right Border of T- DNA of *Agrobacterium tumefaciens*, 35Sde – CaMV35S promoter with duplicated enhancer, *bar* – coding sequence of the herbicide resistance-conferring *bar* gene, 35SpA – polyadenylation signal of Cauliflower Mosaic Virus, *ocspA* – polyadenylation signal of *octopine synthase* gene, *bs(mod)* – modified *barstar* gene, *bs(wt)* – wild type *barstar* gene, TA29(870) – 870bp fragment of the tapetum-specific TA29 promoter, A9(1500) – 1.5kb fragment of the A9 promoter.

Fig. 6 : is a schematic representation of another DNA construct (Construct A) of the invention depicting various components and locations of restriction enzyme sites. Abbreviations: LB – Left Border of T- DNA of *Agrobacterium tumefaciens*, RB – Right Border of T-DNA of *Agrobacterium tumefaciens*, 35S – CaMV35S promoter with single enhancer, *hpt* – coding sequence of the hygromycin resistance conferring *hpt* gene, 35SpA – polyadenylation signal of Cauliflower Mosaic Virus, *ocspA* – polyadenylation signal of

*octopine synthase* gene, *bs(mod)* – modified *barstar* gene, *bs(wt)* – wild type *barstar* gene, TA29(870) – 870bp fragment of the tapetum-specific TA29 promoter, A9(1500) – 1.5kb fragment of the A9 promoter.

Fig. 7 : is a photograph depicting Southern hybridization analysis from the Right  
 5 Border flank of a representative population of transgenic plants generated using Construct A of the invention. Copy number on the right border flank was determined by probing *HincII* digests of genomic DNA with the coding sequence of the wild type *barstar* gene. Lanes 1-13 represent transgenic *barstar* lines while lane 14 represents an untransformed control plant. The numbers along the left margin of the figure represent a non-linear scale,  
 10 in kilobase pairs (kb), of DNA fragment sizes.

To discuss in further detail, the invention involves transformation of a plant cell using the DNA constructs (I and A) of the invention comprising two versions of the fertility restorer gene (*barstar*) – one representing the wild type sequence and the other, a modified sequence for expression in dicotyledonous crop plants – each under independent  
 15 transcriptional control of regulatory elements with overlapping expression pattern in male reproductive tissues and a marker gene (*bar* or *hpt*) under the control of a strong constitutive promoter (CaMV35Sde or CaMV35S) as described in the foregoing sections.

A schematic representation of the Constructs I and A are shown in Figure 5 and Figure 6 respectively of the accompanying drawings. The said transformation can be  
 20 performed employing various vectors and well established procedures such as PEG-mediated direct gene transfer and Electroporation (Bilang et al 1994, *In Plant Mol. Biol. Manual*, Gelvin and Schilperoot, Eds. Kluwer Academic Publishers), Particle Bombardment (Christou 1994, *Plant Mol. Biol. Manual*, Gelvin and Schilperoot, Eds. Kluwer Academic Publishers), etc. The applicants recommend the use of Ti-plasmid  
 25 vectors for genetic transformation of plants because of better control over the process of gene transfer and selection of transgenic plants. In cases of *Agrobacterium*-mediated transformation using disarmed Ti-plasmid vectors, the marker genes are preferably located towards the left border sequence of T-DNA so as to enable selection of transformants carrying the complete T-DNA.

30 The method outlined in the present application can be used, with appropriate modifications, to enhance tissue-specific as well as constitutive expression of transgenes. In the latter case, the inventors recommend the use of different constitutive promoters to express the wild type and modified versions of the candidate gene(s) to circumvent gene silencing due to promoter homology. Several such constitutive promoters viz., CaMV35S,

MAS, OCS, nos, FMV and MMV have been reported in literature. Using the strategy outlined in the present application, the restorer gene, the tissue-specific promoters, the marker gene and the promoter expressing the marker gene can also be used to restore male fertility in a variety of plants wherein transgenic male sterile plants have been generated using molecular methods.

The DNA constructs (I and A) developed in this application are transformed into *Brassica juncea* by *Agrobacterium*-mediated transformation. Transformed shoots are selected on 10mg/L of phosphinothricin (selective agent for the *bar* gene) and/or 20mg/L of hygromycin. Transformed shoots obtained above are transferred to a rooting medium to obtain complete transformed plants, which are subsequently transferred to field conditions and grown to maturity for analysis of their male fertility/sterility status.

Male sterile plants used for restoration studies are characterized by the presence of rudimentary and flattened anthers in flowers with complete absence of pollen production. In contrast, in flowers of untransformed plants and those of restored plants, anthers show pollen production and pollen viability is found to be in the range of 90 – 100%. The frequency of obtaining restored (fertile) plants (on transformation of male sterile lines) and their pollen viability were used as one of the experimental tools to analyze the efficiency of the (fertility restoration) method described in the present application. As described in the foregoing sections, on transforming a male sterile plant with a construct containing two different versions of the restorer gene sequence – one represented by the naturally occurring wild type sequence and the other modified for expression in dicotyledonous crop plants – each under independent transcriptional control of two overlapping tapetum-specific promoters, the frequency of male sterile plants obtained (89.2%) as well as the percent of fertile plants with pollen viability >90% (65%) were significantly higher as compared to those obtained using constructs containing a single copy or two identical copies of the restorer gene. To further confirm the above observations, transgenic *barstar* plants were generated using Construct I of the present invention and single copy *barstar* lines were crossed with male sterile *barnase* lines to analyze restoration of fertility in F1 progeny. It was found in the above study that the construct based on use of two different versions of the restorer gene sequence provides restorer lines at a much greater frequency (~57%) than constructs containing either two identical copies of the restorer gene (43%) or a single copy of the wild type gene sequence (12%).

The present invention provides a successful method for enhancing transgene expression concomitantly reducing susceptibility of the transgene to post-transcriptional

gene silencing. It is important to note that the strategy outlined in the present application can be used with appropriate modifications to enhance both constitutive as well as tissue-specific expression of other candidate genes. Further, it also provides an effective method for the development of efficient fertility restorer lines in crop plants for hybrid seed  
 5 production using molecular methodologies.

## EXAMPLES

### Example 1

10 Construction of modified *barstar* gene:

The 273bp modified sequence of the *barstar* gene was constructed by Recursive PCR (Dillon and Rosen 1990, Biotechniques 9:298-299). The modified sequence was assembled in a first PCR reaction (PCR-I) using four long oligos (L1-L4; with an overlap of 20bp between adjacent oligos) followed by amplification in a second PCR (PCR-II) of  
 15 the assembled sequence using two shorter, flanking primers (S1 and S2). Sequences of the oligos and primers used are given below.

Oligo L1:

5'- ATG AAG AAG GCT GTG ATC AAT GGA GAA CAA ATC AGA TCT ATC  
 TCA GAC CTT CAT CAA ACT TTG AAG AAG GAG CTT GCT CTT CCT G - 3'  
 20 - (Seq. ID # 5)

Oligo L2 :

5' - AAG AGG GTA CTC AAC CCA TCC AGT AAG ACA ATC CCA CAA AGC  
 GTC CAA GTT CTC ACC ATA GTA CTC AGG AAG AGC AAG CTC CTT  
 C - 3' - (Seq. ID # 6)

25 Oligo L3 :

5' - GGA TGG GTT GAG TAC CCT CTT GTT TTG GAA TGG AGG CAA TTC GAG  
 CAA TCT AAG CAA CTT ACT GAG AAT GGA GCT GAG AGC GTT CT - 3' -  
 (Seq. ID # 7)

Oligo L4 :

30 5' - TTA AGA AAG AAT GAT AGT GAT GTC ACA TCC TTC AGC CTT AGC TTC  
 TCT AAA CAC TTG AAG AAC GCT CTC AGC TCC ATT - 3'...(Seq. ID # 8)

Forward primer (S1):

5' - GGC TCG AGC CAC CAT GAA GAA GGC TGT GAT - 3' - (Seq. ID # 9)

Reverse primer (S2):

5' - CTA GTC TAG ATT AAG AAA GAA TGA TAG TG - 3' - (Seq. ID # 10)

Optimal reaction conditions used for the assembly of the modified sequence (PCR-I) consisted of 500ng each of oligos L1-L4, 200μM each dNTP, 2.5U of Amplitaq DNA polymerase and the supplied buffer at 1X concentration in a final reaction volume of 100μl. Cycling parameters were as follows: initial denaturation at 94°C, 4min followed by 4 cycles each of 94°C, 1min; 55°C, 1min and 72°C, 1minute; final extension at 72°C, 5min. A 1μl aliquot of PCR-I was taken as the template for the second round of PCR (PCR-II) using the shorter flanking primers, S1 and S2. Reaction conditions for PCR-II consisted of 500ng each of primers S1 and S2, 200μM each dNTP, 2.5U of Native Pfu DNA polymerase and the supplied buffer at 1X concentration in a final reaction volume of 100μl. Cycling parameters were as follows: initial denaturation at 94°C, 4min followed by 25 cycles each of 94°C, 1min; 55°C, 1min and 72°C, 1minute.

The amplified product was purified and cloned into pCR-Script SK+ at the SrfI site using the pCR-Script SK+ Cloning Kit (Stratagene). Positive clones from two independent amplification reactions were sequenced and were found to contain random mutations in the modified gene sequence. One of the clones containing three single bp additions was subjected to PCR-mediated site-directed mutagenesis to eliminate the point mutations incorporated during gene synthesis. While one of the point mutations was removed by a SOE-ing reaction (Horton 1993, *In PCR Protocols : Current Methods and Applications* Ed. Bruce A. White, Humana Press), the other two mutations were removed in a simultaneous double mutagenesis reaction by Inverse PCR (Imai *et al* 1991, *Nucl. Acids Res.* 19:2785).

For the SOE-ing reaction, two overlapping mutagenizing primers (M1 and M2) each with a single base deletion (corresponding to the added single base mutation) were used in conjunction with the shorter primers, S1 and S2, used earlier for gene synthesis. Sequences of the mutagenic primers are given below:

M1 : 5' - TCC CAC AAA GCG TCC AA - 3' - (Seq. ID # 11)

M2 : 5' - TTG GAC GCT TTG TGG GA - 3' - (Seq. ID # 12)

Two independent PCRs (PCR-Is) were performed using primer sets S1-M1 and S2-M2. Optimized reaction conditions for each amplification consisted of 100ng of template plasmid, 100pmol of each primer, 200μM each dNTP, 2.5U of native Pfu DNA polymerase, DMSO to a final concentration of 5% and the supplied buffer at 1X concentration in a final reaction volume of 100μl. Cycling parameters were as follows: initial denaturation at 94°C, 2min followed by 28 cycles each of 94°C, 30sec; 45°C, 30sec; 72°C, 30seconds; final extension at 72°C, 5min. The amplified fragments were gel-



purified and 100ng of each eluted PCR-I product was taken for PCR-II (SOE-ing reaction) with native Pfu DNA polymerase using primers S1 and S2. Optimized reaction conditions for the SOE-ing reaction consisted of 50pmol of each primer, 200μM each dNTP, DMSO to a final concentration of 5%, 100ng of each amplified eluted product from PCR-I, 1.25U  
5 of native Pfu DNA polymerase and the supplied buffer at 1X concentration in a final reaction volume of 50μl. Cycling parameters were as follows: initial denaturation at 94°C for 2min followed by 25 cycles each of 94°C, 30sec; 50°C, 30sec; 72°C, 30seconds; final extension at 72°C, 2min. The amplified product was gel-purified and cloned into pCR-Script SK+ at the SrfI site using the pCR-Script SK+ cloning kit. Four independent clones  
10 were sequenced and all were found to have incorporated the desired alteration at the desired position.

For Inverse PCR-mediated mutagenesis, two mutagenizing primers (M3 and M4) were designed such that they were divergent from either side of one of the bases to be deleted; the other deletion to be introduced was incorporated into one of the primers (M4).  
15 Sequences of the mutagenizing primers are as follows:  
M3 : 5' - CAA GAG GGT ACT CAA CCC ATC CAG TAA GA - 3' - (Seq. ID # 13)  
M4 : 5' - TTT TGG AAT GGA GGC AAT TCG AGC AAT CTA A - 3' - (Seq. ID # 14)

The above primers were used to amplify the entire plasmid (pCR Script SK+ harboring the cloned, synthetic *barstar* gene) of 3.2kb. Template DNA used for Inverse PCR was subjected to serial 1:10 dilutions to minimize carryover contamination of template plasmids into the amplified mutagenized product. Reaction conditions consisted of ~2ng of template DNA, 100pmol of each primer, 200μM each dNTP, 5U of Taq DNA polymerase and the supplied buffer at 1X concentration in a final reaction volume of  
20 100μl. Cycling parameters were as follows: initial denaturation at 94°C for 4min followed by 20 cycles each of 95°C, 1min; 72°C, 30sec; 72°C, 3min 20seconds; final extension at 72°C for 1min was also given. PCR products from the above amplification reactions were gel-purified, blunt ended, phosphorylated and self-ligated. Following transformation into *E.coli*, clones were sequenced to confirm incorporation of the desired mutation.

30 Following incorporation of the desired modifications in the *barstar* gene, the synthetic gene was mobilized into appropriate cloning vectors for plant transformation as described below.

**Example 2****Development of DNA Constructs:**

The wild type and modified sequences of the *barstar* gene were PCR amplified using the following primers to incorporate the restriction sites BspHI and XbaI at their 5' and 3' ends respectively to facilitate further sub-clonings.

Primers for the wild type *barstar* gene:

FP : 5' - CCT CAT GAA AAA AGC AGT CAT TAA C - 3' - (Seq. ID # 15)

BspHI

10 RP : 5' - GGT CTA GAT TAA GAA AGT ATG ATG GT - 3' - (Seq. ID # 16)

XbaI

Primers for the modified *barstar* gene:

FP : 5' - CCT CAT GAA GAA GGC TGT GAT CAA - 3' - (Seq. ID # 17)

BspHI

15 RP : 5' - CTA GTC TAG ATT AAG AAA GAA TGA TAG TG - 3' - (Seq. ID # 18)

XbaI

Following amplification, the PCR products were cloned in pCR Script SK+ at the SrfI site to generate the plasmids pCR Script SK+ *barstar*(wt) and pCR Script SK+ *barstar*(mod) and sequenced to verify the fidelity of amplification.

20

**Development of Construct I:**

- a) The *bar* gene, used as a plant selectable marker, was fused at its 5' end with the CaMV35S double enhancer promoter (CaMV35Sde) and to a suitable transcription termination/polyadenylation signal at its 3' end.
- 25 b) The wild type *barstar* gene was fused at its 5' end with an 870bp fragment of the tapetum-specific TA29 promoter and with a suitable transcription termination/polyadenylation signal at its 3' end.
- c) The modified *barstar* gene was fused at its 5' end with the tapetum-specific A9 promoter and at its 3' end with a suitable transcription termination/polyadenylation signal.
- 30 d) The components described in (a) to (c) were mobilized into the binary vector pPZP200 as described below:

The CaMV35Sde-*bar*-ocspA transcription unit was isolated as a PstI-HindIII fragment, blunt ended using Pfu polymerase and cloned at the SmaI site of pPZP200

followed by cloning of the A9(1500)-*barstar*(fm)-pA transcription unit as a PstI fragment in the PstI site of the same. The TA29(870)-*barstar*(wt)-pA transcription unit was isolated as a HindIII fragment and cloned at the HindIII site of above binary vector to obtain the final construct.

#### 5 Development of Construct A:

- a) The *hpt* gene, used as a plant selectable marker, was fused at its 5' end with the CaMV35S promoter (35S) and to a suitable transcription termination/polyadenylation signal at its 3' end.
- b) The wild type *barstar* gene was fused at its 5' end with an 870bp fragment of the tapetum-specific TA29 promoter and with a suitable transcription termination/polyadenylation signal at its 3' end.
- c) The modified *barstar* gene was fused at its 5' end with the tapetum-specific A9 promoter and at its 3' end with a suitable transcription termination/polyadenylation signal.
- d) The components described in (a) to (c) were mobilized into the binary vector pPZP200 as described below:

The A9(1500)-*barstar*(fm)-pA transcription unit was isolated as a PstI fragment and cloned in the PstI site of the binary vector pPZP200 followed by cloning of the TA29(870)-*barstar*(wt)-pA transcription unit as a HindIII fragment at the HindIII site of above binary vector. The 35S-*hpt*-pA transcription unit was isolated as a SacI-SalI fragment and cloned in the corresponding sites of the above binary to develop the final construct.

The final transformation vectors were mobilized into *Agrobacterium tumefaciens* strain GV3101 by electroporation using the BioRad Gene Pulser according to manufacturers' instructions.

#### Example 3

##### Seed Sterilization and Germination:

Seeds were surface sterilized by treatment in diluted Teepol solution (laboratory grade surfactant) for 10 minutes followed by washing under running water for 30 minutes. Seeds were subsequently treated with 70% ethanol for two minutes under sterile conditions and rinsed twice with sterile distilled water. Further, seeds were treated with 0.05% mercuric chloride for 10 minutes followed by a treatment with 1% sodium hypochlorite for 9 minutes. Following each treatment, seeds were rinsed thoroughly with

sterile distilled water. Surface sterilized seeds were germinated on hormone-free full strength MS medium (Murashige and Skoog 1962, *Physiol. Plant.* 15:473-493). Seeds were germinated in glass tubes covered with a cotton plug with four seeds in each tube. The seeds were kept in dark for two days and then maintained under light (Philips cool-white fluorescent lamps, 2000lux, 16-h light/8-h dark cycle). Temperature in the culture room was maintained at  $23\pm 1^{\circ}\text{C}$ .

#### Example 4

##### *Agrobacterium*-mediated genetic transformation of *Brassica juncea*:

- 10 Transformation of *Brassica juncea* was performed according to the protocol described by Bade and Damn (1995, *In Gene Transfer to Plants*, Potrykus and Spangenberg, Eds. Springer Lab Manual) with further modifications. A single bacterial colony harboring the desired construct was inoculated in selective medium containing appropriate antibiotics and grown to saturation. A secondary culture was initiated from the
- 15 saturated primary culture and grown for 3 hours at  $28^{\circ}\text{C}$  in non-selective medium. The bacterial cells were subsequently harvested by centrifugation at 5000rpm for 15min at  $22^{\circ}\text{C}$  and resuspended to a final  $\text{OD}_{600}$  (optical density at 600nm wavelength) of 0.3 in MS medium containing plant growth hormones, 1mg/ml BAP (6-benzylaminopurine) and 1mg/ml NAA ( $\alpha$ -naphthalene acetic acid). This suspension was used for infection of
- 20 explants. Hypocotyls of 5-day old seedlings of *Brassica juncea* (germinated on an appropriate medium as described above in Example 3) were cut into 5mm long segments and precultured in aforementioned medium (hereinafter referred to as MSN1B1) for 18 hours at  $22^{\circ}\text{C}$  with mild shaking at 110rpm. Following preculture, the medium was decanted and explants were infected for 30 minutes with the bacterial suspension prepared
- 25 as described earlier. The bacterial suspension was then replaced with MSN1B1 medium and the hypocotyl explants were cultured for 12-14 hours under similar conditions as described above. The explants were subsequently washed with MSN1B1 containing 200mg/L of the bacteriostatic agent augmentin (to restrict *Agrobacterium* growth) and plated on appropriate selective media.
- 30 For transformation of *Brassica juncea* using Construct I (containing the *bar* gene as a selectable marker), explants were plated on MSN1B1 containing 10mg/L of phosphinothricin. Regenerated shoots were transferred after six weeks to MS medium containing 2mg/L IBA (Indole 3-Butyric Acid) for rooting and were maintained as nodal cultures until transplantation.

In case of transformations using Construct A (containing the *hpt* gene as a selectable marker), hypocotyl explants derived from backcrossed seeds of the male sterile *barnase* line *bn* 3.6 (containing the *bar* gene as a selectable marker) were used. Since the above *barnase* line contains a single copy of the T-DNA insert, the backcrossed progeny would segregate for phosphinothricin resistance (associated with male sterility) and sensitivity in a 1:1 ratio. Following infection, co-cultivation and washing of explants as described above, explants were plated initially on MSN1B1 medium containing 20 mg/L of the selective agent hygromycin. Regenerating explants were transferred after 25-30 days to MSN1B1 medium containing both the selective agents, hygromycin (20 mg/ml) and phosphinothricin (10 mg/ml). Putative transformed shoots were transferred for rooting on MS medium supplemented with 2mg/ml IBA, hygromycin (20 mg/L) and phosphinothricin (10 mg/L). Transgenic plants were maintained as nodal cultures until transplantation.

#### Example 5

Analysis of male fertility/sterility status and pollen viability

Male fertile/sterile plants were identified on the basis of morphological observations of anthers and presence/absence of pollen production. Two to three inflorescence axes from each plant (each inflorescence bearing 10-15 unopened buds) were covered with a pollination bag to test for the formation of selfed seeds. Absence of selfed seeds was taken as a confirmation of male sterility. In plants wherein pollen formation was observed, pollen viability was tested by fluorescein diacetate (FDA) staining according to the protocol described earlier (Heslop-Harrison *et al* 1984, Theor. Appl. Genet. 67:367-375).

#### Example 6

Molecular characterization of transgenic plants:

(a) Isolation of total DNA:

Total DNA was isolated from fully expanded leaves of transgenic plants (growing in the field under containment conditions) and the untransformed parents following Rogers and Bendich (1994, *In Plant Mol. Biol. Manual*, Gelvin and Schilperoot, Eds. Kluwer Academic Publishers). One gram of leaf tissue was finely powdered in liquid nitrogen and homogenized in 5ml extraction buffer containing 100mM Tris-HCl pH 8.0, 20mM Sodium EDTA (ethylene diamine tetraacetic acid) pH 8.0, 1.4M Sodium Chloride, 1%

PVP40 (polyvinyl pyrrolidone 40) and 2% CTAB (Cetyltrimethyl Ammonium Bromide). The above material was incubated at 65°C with occasional shaking followed by extraction with an equal volume of chloroform:isoamylalcohol. To the supernatant obtained above, 1.25ml of 10% CTAB solution was added followed by extraction with an equal volume  
5 Chloroform : isoamylalcohol. Genomic DNA was precipitated from the above supernatant by addition of three volumes of precipitation buffer (50mM Tris-HCl pH 8.0, 10mM Sodium EDTA and 1% CTAB) followed by incubation at room temperature for 30 minutes. The pellet obtained was dissolved in 500µl of buffer containing 10mM Tris-HCl pH 8.0, 1mM Sodium EDTA and 1M NaCl. Undissolved  
10 impurities were removed from the above sample by centrifugation followed by precipitation of dissolved DNA using 100% ethanol. The DNA pellet was washed with 70% ethanol and finally dissolved in an appropriate amount of sterile distilled water.

(b) Protocols for Restriction Digestion and Southern Hybridization:

15 Ten micrograms of total genomic DNA were digested overnight with appropriate restriction enzyme(s) in a 70µl reaction volume containing 40 units of the restriction enzyme under conditions as recommended by the manufacturers. The digested DNA was electrophoresed on a 0.8% agarose gel at 1.75V/cm for 18-20 hours. Following electrophoresis, the restricted DNA was transferred onto a nylon membrane (Hybond  
20 N+, Amersham) by capillary action (for 12-14 hours), air-dried for 45min and cross-linked in a UV-Crosslinker (Amersham) at  $7 \times 10^4$  joules/cm<sup>2</sup>. Probes used for hybridization were labeled with  $\alpha$ -<sup>32</sup>PdCTP or  $\alpha$ -<sup>32</sup>PdATP by a random priming method using the Megaprime DNA Labeling System (Amersham Pharmacia Biotech). Following hybridization, blots were washed twice in 2X SSC buffer (30mM sodium chloride and  
25 0.3M sodium citrate, pH 7.6) at 25°C for 15min each, followed by one wash at 65°C for 15 minutes. Stringent washes (wherever necessary) were performed in 0.2X SSC, 0.1% SDS at 65°C for 15 minutes. The blots were subsequently covered with saran wrap and exposed to X-ray films (Kodak) for 12-24 hours at minus 80°C. Prior to reprobing, blots were deprobed for 40min in 0.4N NaOH at 42°C followed by treatment with a  
30 neutralization solution (0.2M tris pH 8.0, 0.1X SSC, 0.5% SDS) for 40 min at 42°C.

(c) Analysis of Copy Number:

Transgenic plants generated in *Brassica juncea* were subjected to Southern hybridization to determine copy number of the T-DNA insert and to identify single copy

plants for analysis and further studies. Genomic DNA isolated from leaves of transgenic plants generated using DNA Constructs I, II and III were digested with the restriction enzyme BamHI while genomic DNA isolated from leaves of fertile transgenic plants generated by retransformation of the male sterile *barnase* line *bn* 3.6 using Constructs A, B and C were initially digested with the enzyme HincII (for copy number determination on the right border flank) followed by digestion of single copy events identified above with XbaI (for determination of copy number on left border flank). Restriction digests were electrophoresed on a 0.8% agarose gel, transferred onto a nylon membrane and cross-linked using a UV-crosslinker. The Southern blots thus generated for Constructs I, II and III were probed with CaMV35Sde promoter for LB and *barstar*(wt) fragment for RB, while those generated using Constructs A, B and C were probed using the EcoRV-HincII fragment of CaMV35S-*hpt*-pA cassette representing partial sequences of the CaMV35S promoter and the *hpt* gene for left border of T-DNA and the right border integration was identified using the *barstar*(wt) fragment. In all cases, probes used represent DNA sequences derived from both sides of the restriction enzyme site(s) used for digestion of genomic DNA so as to enable determination of copy number on both flanks of the T-DNA.

#### ADVANTAGES

- 1) The present invention provides an effective method for enhancing transgene expression in heterologous systems.
- 2) The present invention provides a novel combination of two sequence versions of a candidate transgene (*barstar*) – one representing the wild type sequence and the other, a modified sequence designed for expression in dicotyledonous crop plants – with two different overlapping tissue-specific promoters (TA29 and A9), a strong constitutive promoter (CaMV35S with double or single enhancer) and a marker gene (*bar* or *hpt*) in a novel construct to enhance transgene expression.
- 3) The present method relies on the simultaneous use of two different (target) gene sequences encoding identical protein products in the same DNA construct to enhance expression levels. Since the two different gene sequences would transcribe two different RNA molecules, they would effectively circumvent post transcriptional gene silencing due to coding sequence homology.

- 4) The present method can be effectively used for enhancing constitutive as well as tissue-specific expression of a transgene.
- 5) The present method provides an effective method for development of efficient restorer lines (by significantly enhancing restorer gene expression) in crop plants for hybrid seed production using molecular methodologies.

5

10

15

20

25

30